

THE OPEN UNIVERSITY OF ISRAEL OPEN MEDIA AND INFORMATION LAB

Rethinking access tools for researching multimodal digital collections

Vered Silber-Varod

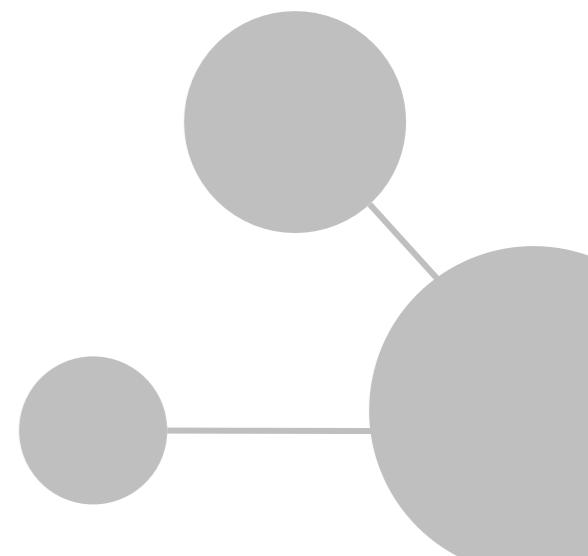
EVA / Minerva 2017

XIVth Annual International Conference for Professionals in Cultural Heritage
14th November, 2017
The Van Leer Jerusalem, Institute



CHALLENGE

- Traditional research methods in the social sciences (surveys, interviews, questionnaires) do not properly address all social phenomena in the digital era.
- While the industry is oriented towards big data, social scientists and students lack skills and methods for using public digital data for research.

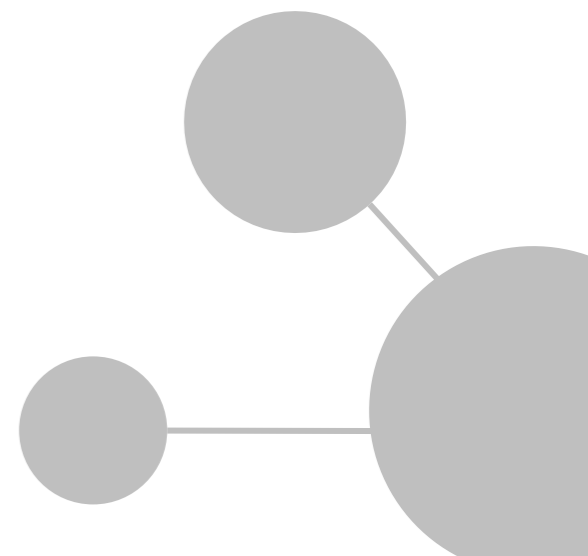


MISSION STATEMENT

- Building open access digital tools for social research, to be used by researchers, students, journalists and activists.
- Making digital research tools accessible.
 - Lowering the entry level to digital research for students and social scientists without prior knowledge in programming.
- Using public digital data for the benefit of research, teaching and society.

Why with NLI?

- Wealth of digital data and collections of tremendous importance for the general public, and in particular for the community of students and researchers.
- The goal of this research is to augment NLI's collections and increase their value to academia and to society in general.



Five-divisions project



 The Open University of Israel



Hebrew Journalism Archives

PI: Prof. Oren Soffer

- Conjoining Historical Digitized and Contemporary Born-Digital Newspaper Archives.

Expected Outputs:

- An advanced search and access interface to news archives.
- A topic modeling analysis tool for viewing their evolution over time.
- A geo-located app that displays the news that were reported about the user's location throughout history.

Israeli Web Archive

PI: Dr. Anat Ben David

- Extracts archived pages from the Internet Archive and reconstructs historical networks of the Web's past
- **Expected outputs:**
 - Advanced search scenarios
 - Image analysis
 - Geo-location analysis of the Israeli Web Archive

Audio/Video to Text

PI: Dr. Anat Lerner and Dr. Vered Silber-Varod

- Providing researchers in the humanities with a vast amount of new types of research data (e.g., dialogues in films, interviews, broadcastings).
- To apply speech-analysis-techniques to these large-scale collections of speech-based audio and video.
- **Expected outputs:**
 - Modeling structure and meaning of different types of audio and video semiotic artifacts .
 - Automatic transcripts.
 - Synchronization with the auditory component.
 - Searchable.

Entity Extraction (NLP)

PI: Dr. Reut Tsarfaty

- Automatic Hebrew text analyzer for recognizing named entities of different types: Person, Location, Organization, etc.
- **Expected outputs:**
 - A statistical, robust and broad-coverage, named entity recognizer, that fits the peculiar structure of Hebrew texts.
 - Tools uniquely modified to support analysis of content in Hebrew and Arabic.

Educational Project: Students' lab

PI: Prof. Yoram Kalman

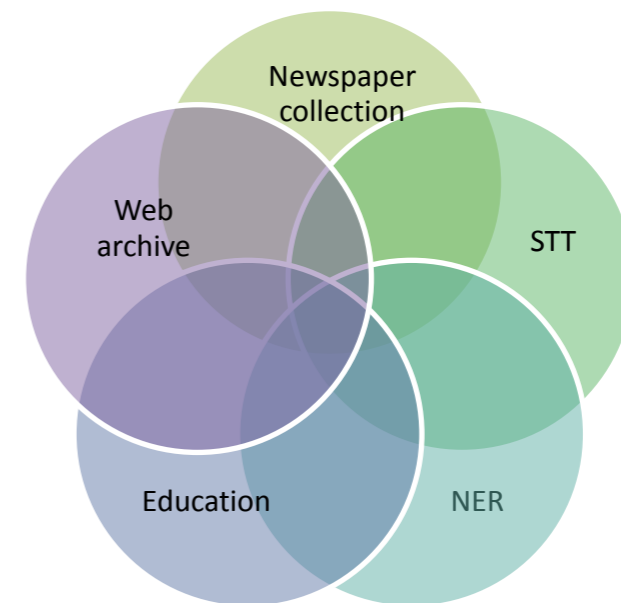
Data mining

Media
monitoring

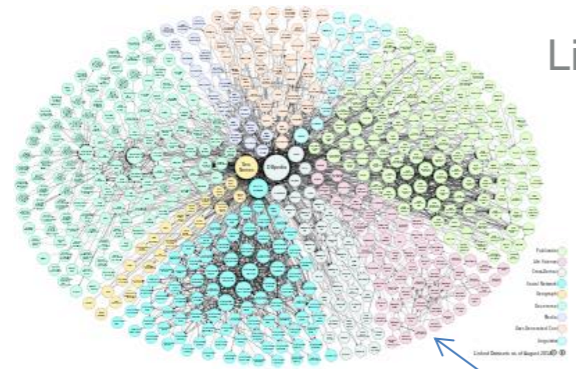
Big data
analytics

Social
network
analysis

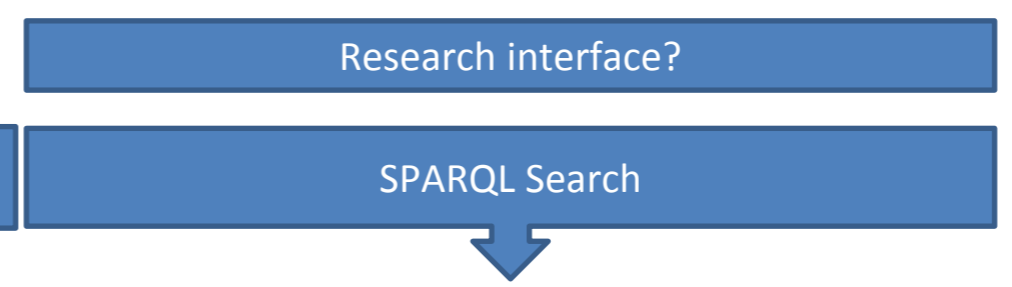
- Multimodal approach of analysis.
- Synthesis of results.
- Results are given per query. Not “per collection”.
- Example: Texts refer to negative sentiments towards PMs between the years 1970-2000.



WWW

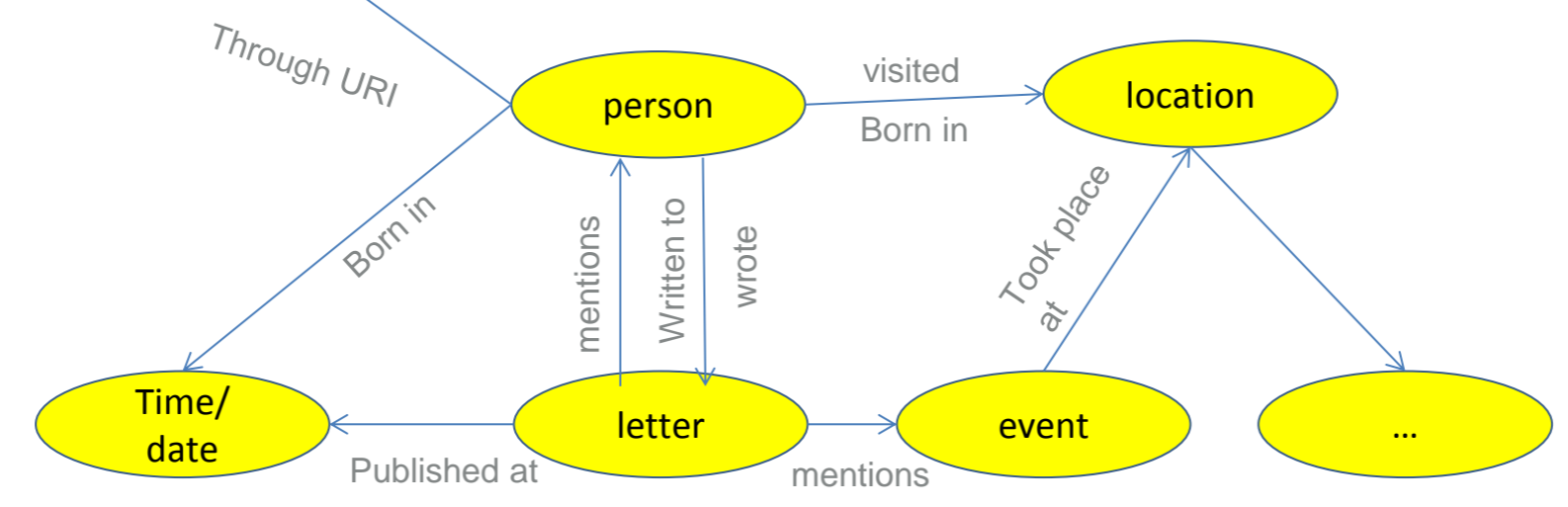


Linked data: e.g., BBC



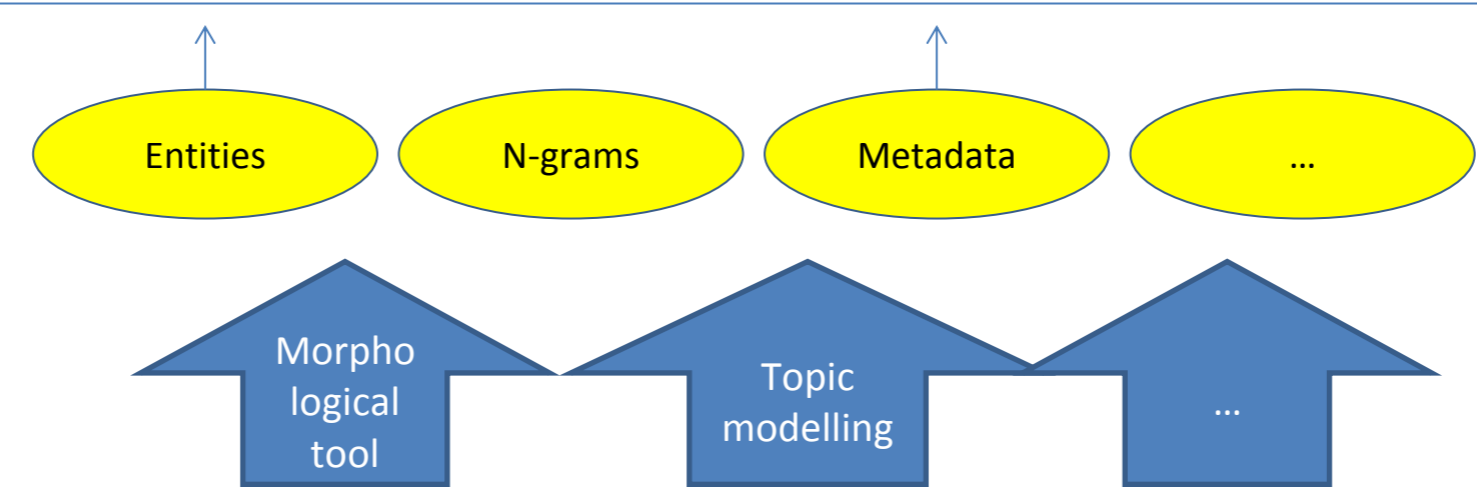
Semantic querying

Lab servers

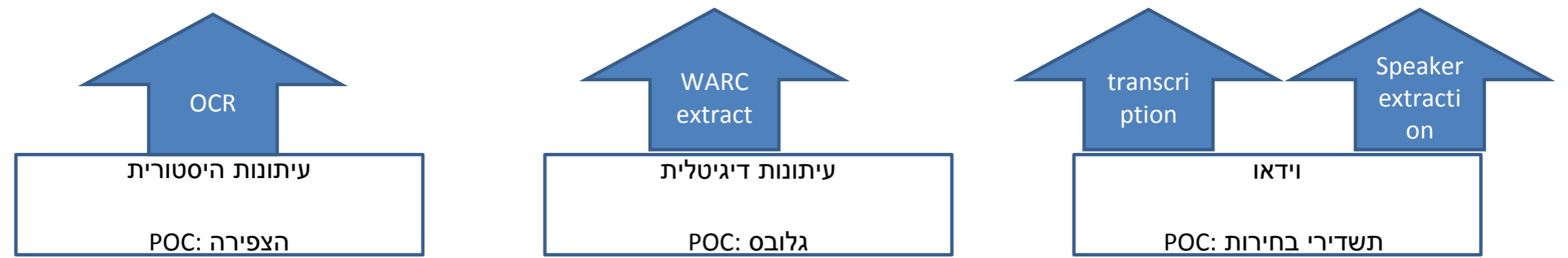


Hebrew text analysis

Lib servers



Raw data



עיתונות היסטורית

POC: הצפירה

עיתונות דיגיטלית

POC: גלובס

וידאו

POC: תשדירי בחירות

Thank you



 The Open University of Israel